Working Paper Series

An intra-group variance based polarization measure

Federico Palacios-González
Rosa García-Fernández

ECINEQ WP 2011 – 188

ECINEQ 2011-188
January 2011

www.ecineq.org

An intra-group variance based polarization measure*

Federico Palacios-González
Rosa García-Fernández†
University of Granada

Abstract

In this paper, it will be shown that polarization not only depends on identification and alienation. There exist additional factors such as the share of power held by each of the groups and the number of groups that should be considered to compute it. Consistent with this idea and with the basic principles of polarization of Esteban and Ray (1994), we define a measure of polarization as the product of three factors: the alienation-identification, the share of power of the groups and the number groups. The first factor is defined following an identification-alienation framework based on the decomposition of the variance. This differs from the expression defined by Esteban and Ray (1994) and Zhang and Kanbur (2001). The second factor values the importance of the middle class as a factor of social stability. This factor, jointly with the third, extends and corrects the approach to polarization of Zhang and Kanbur (2001). A comparison of the defined measure is made with the main expressions existing in the studies about income polarization. For this, data derived from the ECHP (1994-2001) for Spanish Households are utilized.

Keywords. Polarization, inequality, conflict, k-means algorithm
JEL Classification: D31,D63

* The authors are grateful to Esteban J.M. for his useful comments in a previous version of the paper. We also thank the comments and suggestions received in the Research Workshop of the Israel Science Foundation on Income Polarization: Measurement, Determinant, Implications. Especially we thank Gordon Anderson for his useful comments and suggestions.
† **Addresses of correspondence**: Department of Quantitative Methods for Economics and Business University of Granada, Campus de Cartuja s/n, Granada, 18071, Spain. Emails: fpalacio@ugr.es, rosamgf@ugr.es.

1. Introduction[1]

There is increasing interest in studying the social changes that contribute to the clustering of the population into homogeneous groups that oppose each other, thus raising the potential for political and social conflicts. This idea is contained within the concept of polarization introduced by Wolfson (1994) and Esteban and Ray (1994) independently, and following different theoretical approaches. The relevance of this topic has aroused great attention resulting in the publication of numerous articles since the mid-nineties until now [see among others Tsui and Wang (2000), Zhang and Kanbur (2001), Duclos et al. (2004), Silber et al. (2007), Zelli and Pittau (2007), Gasparini et al. (2008) and Hussain (2009)]. However, there are in our opinion, several issues related to the measurement of polarization that should be studied in greater depth. With this aim, in this paper an index of polarization is developed, based on the following basic features that, according to Esteban and Ray (1994), such a measure must present: a) there must be a high degree of homogeneity within each group; b) there must be a high degree of heterogeneity across groups and c) there must be a small number of significantly-sized groups. In compliance with these characteristics, the proposed measure depends on three factors: the homogeneity or identification within group and the heterogeneity or alienation between groups, the share of power or social weight of the groups and the number of significantly sized groups. To quantify the effects that the previous factors have on polarization, we define three indices and the product of them provides a straightforward tool to compute an index of income polarization.

The new index is defined considering the alienation-identification framework to polarization put forward by Esteban and Ray (1994, ER henceforth) and Zhang and Kanbur (2001, ZK henceforth) and consequently, raises the following question: what does the new measure add to those proposed by the mentioned authors? In first place, the developed index introduces important changes to compute alienation and identification. We presume that heterogeneity or alienation is linked to the distance between the mean incomes of the groups. Additionally, we consider that an individual feels identified with the group to which he or she belongs, when his or her income is closer to the average income of the group. Focusing on the decomposition of the variance, we assume that alienation is proportional to the inter-groups variance, and identification is inversely proportional to the intra-group variance. The ratio of the inter-

[1] The authors are grateful to Esteban J.M. for his useful comments in a previous version of the paper. We also thank the comments and suggestions received in the Research Workshop of the Israel Science Foundation on Income Polarization: Measurement, Determinant, Implications. Especially we thank Gordon Anderson for his useful comments and suggestions.

groups heterogeneity to the intra-group homogeneity [2] will be utilized to define an index, I_{ia},that quantifies the contribution of identification and alienation to polarization. The use of the variance is justified by statistical reasons as well as by the concept of polarization. From a statistical point of view, the intra-group and the inter-groups variances are the most appropriate approaches to evaluate the homogeneity within a group, and the heterogeneity across groups respectively, when the representative magnitude of each group is the mean of the variable of interest, which is in our case the mean income (see among others Fisher, 1958). Nonetheless, most of the publications related to polarization use concentration as a measure of homogeneity. This is because concentration is used as a synonym of lack of dispersion. However, the former refers to the way in which total income is distributed among individuals, and the latter is related to the degree of homogeneity of the values of the statistical variable. Although there is a correspondence between equidistribution or null concentration and null dispersion, maximum concentration is not associated with either maximum dispersion or with null dispersion. Indeed, both characteristics are not totally opposite, and hence, the contrary of dispersion is homogeneity but not concentration. Therefore, homogeneity and concentration are not equivalent concepts and the measures of concentration should not be used to compute dispersion and vice versa (see for instance Hermoso and Bastida, 2000). In addition, we would like to emphasize that the concept of polarization should be understood in terms of social cohesion and therefore, the measures of polarization, unlike the inequality indices, should not be viewed in a welfare context[3]. Consequently, the expressions used to quantify alienation and identification should not be implicitly or explicitly based on normative measures of concentration such as the Gini or the Theil indices. To illustrate this point we can imagine a society with a high level of welfare in which there are two ethnic groups with competing interests. Even if this society enjoys a high level of welfare, the existence of two well differentiated groups may be a source of social conflict and instability. For this reason, we think that positive measures, such as the variance, which make no explicit use of any concept of social welfare, are more appropriate to compute alienation and identification. On this point the proposed index differs from the measures defined by Wolfson (1994), Zhang and Kanbur (2001) and Silber et al (2007), which use normative inequality indices to compute within group homogeneity and between groups heterogeneity.

[2] As we will observe in section two, this ratio is close to the measure of polarization of Zhang and Kanbur (2001). If we normalized the latter measure using the decomposition property of the Theil index, we obtain an expression that resembles our index.

[3] We thank Daniel Gottlieb for his comments related to this question.

To evaluate the contribution of the share of power[4] of the groups to polarization we build a second index, I_m, taking into consideration the distance between the distribution of the size of the groups and the distribution of maximum polarization. This index captures the effect that the location of significant sized groups around the extremes of the income distribution has on polarization. In particular, the movements of individuals from the middle to the top and to the bottom of the distribution are recorded by this index as an increase in polarization. The following example explains why I_m should be introduced in our measure. Let us consider three income levels 20, 40 and 60 with population shares $(0.1,\quad 0.8,\quad 0.1)$ and mean equal to 40. Suppose that the middle class disappears, all the rest remain the same, so that the new population shares are $(0.5,\quad 0,\quad 0.5)$. In both distributions the groups are homogeneous (the intra-group variance is zero) and consequently our alienation-identification function reaches the maximum. Nonetheless, the degree of confrontation is greater in the second distribution since, the higher the share of power held by each of the groups, the greater their social influence and the potential to initiate conflicts. The introduction of I_m in the measure allows us to compute this change in the share of power as an increase in polarization. The measure of ZK fails in this aspect; indeed it tends to infinite in both situations since the intra-group inequality is equal to zero. The expression of ER performs well because the identification component of this measure is defined as a function that depends uniquely on the size of the groups, and a parameter that indicates the degree of sensitivity to polarization[5].

The contribution of the number of groups to polarization is evaluated by means of a third index, I_g, decreasing with the number of groups, so that the smaller the number of groups is, the higher this factor's contribution to polarization is. The introduction of this index in the measure compensates for the effect that the increasing number of groups has on the intra-group variance, and hence on the contribution of the identification-alienation factor to polarization. Empirically, it is observed that the measure of ZK fails on this point, in such a way that the higher the number of groups, the greater the recorded polarization is.

As we pointed out at the beginning of this section, the product of the three factors of polarization, quantified by the indices I_{ia}, I_m and I_g, provides an index which is a non-decreasing function of the alienation-identification, the distribution of the share of power of the groups and the number of groups. The new index takes values over the interval [0, 1] and

[4] We assume that, at least in a democratic context, the bigger the size of the group, the greater the share of power. Nonetheless, alternative approaches to evaluate the share of power of the groups can be considered without affecting the methodology of calculating the measure.

[5] However, in accordance with the basic features or polarization of ER, we have preferred to use an identification function that depends on dispersion of the income groups.

can be interpreted as a percentage that shows the degree of polarization. This gives it an advantage over the expressions of ER and ZK. The measure of ER is not dimensionless and is not normalized on the interval [0, 1]. Although Esteban and Ray (1994) made an attempt to normalize their measure using log income and replacing the population weights by the population frequencies, the fact is, that this measure can take values higher that one as is shown in the following example. Let us assume that the individuals of a country are bunched into three groups with annual log income levels 5, 10, and 25 and with population shares $(0.1,\ \ 0.8,\ \ 0.1)$. For these data the measure of ER is equal to 1.17[6]. However, due to the fact that scale units of polarization are undefined, since there is not an established standard of measurement, we cannot say that 1.17 means 1.17 times the polarization of a basic element chosen as reference in the standard of measurement of the ER expression. Indeed, for log income the measure of ER does not depend on the income unit and takes values closer to those obtained in the measurement of inequality. Despite the transformation of the data this expression takes values above one and consequently the results cannot be interpreted in terms of percentages. The expression of ZK is not normalized and tends to infinite when the intra-group inequality is equal to zero. Hence, this measure cannot say anything about polarization when everybody in each group possesses an income equal to the mean of the groups.

To calculate the proposed measure, it is necessary to determine the number of groups and their location. Esteban (2002) left the determination of the number of groups to the analyst, and endogenously established the size of the groups, minimizing income inequality within group. In this paper, following this author we compute polarization for two and three groups, but we determine the support interval of each income group applying the k-means algorithm (McQueen, 1967). In this way, for any predetermined value of k, we obtain the k income groups that minimize the intra-group variance and consequently, the inter-groups variance is maximized. Thus, we obtain the k income groups of maximum alienation-identification.

The measure proposed has been applied to empirical data from the European Community Household Panel (1994-2001) for Spanish households. To establish comparisons and enrich the empirical analysis, we calculate the measures of Esteban and Ray (1994), Zhang and Kanbur (2001), Wolfson (1994) and Tsui and Wang (2001) for two groups and the measures of Esteban and Ray (1994) and Zhang and Kanbur (2001) for three groups. In addition we present the inequality indices of Gini and Theil for contrast.

[6] The expression of the measure of ER is in the Appendix. This value has been obtained for $\alpha = 1$.

This paper is organized as follows. Section 2 introduces the new measure of polarization. Section 3 describes the empirical application and Section 4 contains final remarks.

2. A new measure of income polarization

In this section, we develop the proposed measure of polarization assuming that polarization will depend on three factors

1. The cohesion within group and the heterogeneity between groups.
2. The share of power or social weight of the groups
3. The number of significantly sized groups

We begin by focusing on our first factor of polarization, that is, the cohesion within group and the heterogeneity between groups. Using identification as a synonym of cohesion, we assume that identification is related to the similarity of the income within group. An individual feels a sense of identification with the group to which he or she belongs when his or her income is closer to the average income of the group. The smaller the distance, the higher the homogeneity within the group. We presume that heterogeneity or alienation is linked to the distance between the mean incomes of the groups. The larger the distance, the higher the alienation that the individuals in a group feel from the other groups. In line with the previous arguments we consider, that a global measure of identification of individuals of a population with individuals of sub-populations to which they belong, should be inversely proportional to the intra-group variance (V_W). In addition, a global measure of alienation felt by individuals that belong to the same group with respect to individuals belonging to the other groups, should be proportional to the variance between groups (V_B). Considering both as polarization factors we can measure the contribution of the global identification to polarization by means of expression

$$I_1 = c_1 \times V_W^{-1}$$

where $c_1$is a constant of proportionality that defines the scale used.

In a similar way, the contribution of the global alienation to polarization can be quantified by:

$$I_2 = c_2 \times V_B.$$

Hence, the adding of both factors to polarization is given by[7]

[7] Observe that the expression I'_{ia} is independent of the monetary units used to measure income.

$$I'_{ia} = I_1 \times I_2 = c \times \frac{V_B}{V_W} \in [0, +\infty]$$

where $c = c_1 \times c_2$.

I'_{ia} can be normalized as follows

$$I_{ia} = \frac{I'_{ia}}{c + I'_{ia}} \in [0,1].$$

It is obvious that[8]

$$I_{ia} = \frac{I'_{ia}}{c + I'_{ia}} = \frac{c \times \frac{V_B}{V_W}}{c + c\frac{V_B}{V_W}} = \frac{V_B}{V_W + V_B}.$$

Taking into consideration that the variance of the overall population (*V*) is partitioned as follows

$$V = V_W + V_B \ (1)$$

we have that

$$I_{ia} = \frac{V_B}{V} = 1 - \frac{V_W}{V}.$$

According to I_{ia} , as income differences within groups diminish, i.e. the sense of identification increases, differences across groups are enlarged and polarization is higher. That is, alienation and identification are not independent quantities. Indeed, they can be considered two sides of the same coin given that they are linked by the variance of the overall population as shown in expression (1).

Below we introduce some concepts that will be used to define the remaining factors of polarization that is, the share of power of the groups and the number of groups.

To evaluate the impact of the share of power of the groups we are going to focus on the distribution of the size of the groups because, we presume that in a democratic context where everyone has a vote, the greater the size of the groups the greater their share of power. However, alternative approaches to evaluate the share of power can be considered without

[8] Note that the normalized index does not depend on the proportionality constant. Indeed, it is independent of the scale used to measure identification and alienation respectively.

affecting the methodology of calculus of the index of polarization. As we shall see, this component acts a factor that values the importance of the middle class as a factor of social stability. The smaller the size of the middle class the greater the contribution of this factor to polarization.

For a constant number of k significant groups, we assume that the distribution that presents a level of polarization higher than any other distribution,p^H, is given by $p^H = \left(\frac{1}{2}, \quad \frac{1}{2}\right)$ for $k = 2$ and by $p^H = \left(\frac{1}{2}, \quad 0, \dots, \quad 0, \quad \frac{1}{2}\right)$ for $k \geq 3$. Let p represent the distribution of the size of the groups. A measure of distance,d, between p^H and p can be defined which is consistent with polarization, in the sense that higher distance involves lower polarization. It is determined as the Euclidean distance[9] between the distributions of the population, p and p^H, that is

$$d = \sum_{j=1}^{k} \left(p_j - p_j^H\right)^2 = \begin{cases} (p_1 - 0.5)^2 + (1 - p_1 - 0.5)^2 = 2(p_1 - 0.5)^2 \text{ for } k = 2 \\ (p_1 - 0.5)^2 + (p_k - 0.5)^2 + \sum_{j=2}^{k-1} p_j^2 \qquad \text{for } k > 2 \end{cases}$$

The measure of distance d is equal to zero if $p_j = p_j^H \; \forall \, j$. The maximum value reached by d is given by

$$c(k) = \begin{cases} \frac{1}{2} \text{ for } k = 2 \\ \frac{3}{2} \text{for } k > 2 \end{cases}.$$

This maximum is reached with the distributions $(1, \quad 0)$ or $(0, \quad 1)$ for $k = 2$ and with the distribution $(0 \quad \dots \quad 1^{j)} \dots \quad 0) \; \forall j$ such that $1 < j < k$.

In line with the approach followed to compute the alienation-identification factor, let us define the expression

$$I'_m = c \times \frac{1}{d} \in \left[\frac{c}{c(k)}, +\infty\right]$$

where c is a constant of proportionality that defines the scale used.

[9] Note that according to the approach to polarization of Esteban and Ray (1994), this factor is defined assuming symmetry.

I'_m decreases with d, that is, the greater the distance from p_j to the maximum polarization distribution, p_j^H, the greater is its contribution to polarization. I'_m can be normalized as follows

$$I_m = \frac{I'_m - \frac{c}{c(k)}}{I'_m + c}.$$

It is evident that

$$I_m = \frac{\frac{c}{d} - \frac{c}{c(k)}}{\frac{c}{d} + c} = \frac{c(k) - d}{(1+d)c(k)} \in [0,1].$$

Substituting $c(k)$ in the previous expression we have

$$I_m = \begin{cases} \frac{1-2d}{(1+d)} & \text{for } k = 2 \\ \frac{3-2d}{3(1+d)} & \text{for } k = 3,4,\ldots \end{cases}.$$

Observe that I_m captures the effect that the clustering of population around the extremes of the income distribution, or equivalently the influence of a shrinking of the middle class has on polarization. Movements of individuals from the middle to the bottom and to the top of the income distribution will involve an increase of I_m and hence of polarization. If the middle class disappears, leading to a new distribution with two equally sized groups located at the extremes, this index reaches the maximum. Focusing on two groups, there will be more likelihood of generating conflict when the groups become more equal in size. Therefore, those equalizing size movements will augment the value of I_m and polarization.

To compute the contribution of the number of groups to polarization we assume that polarization decreases with the number of significantly sized groups for $k \geq 2$ and it is non-existent for $k = 1$. According to these assumptions we define an index, $I_g(k)$, that represents the contribution of the number of groups to polarization. It is given by

$$I_g(k) = \begin{cases} 0 & k = 1 \\ \frac{2}{k} & k \geq 2 \end{cases}.$$

The introduction of this index into the measure compensates the effect that the increasing number of groups has on the intra-group variance, and hence on the contribution of the identification-alienation factor to polarization. In other words, as the number of groups increase the intra-group variance diminishes enhancing I_{ia}. By introducing I_g this effect is

counterbalanced. This factor extends and corrects the approach of ZK which empirically fails on this point, in such a way that the higher the number of groups, the greater the recorded polarization continually is.

Given that I_{ia} , I_m and I_g are polarization factors that take values over the interval $[0,1]$, their product will provide a normalized and non dimensional index of polarization, that is

$$PG = I_{ia} \times I_m \times I_g \ \in [0,1]$$

Multiplied by 100 the result can be interpreted as a percentage of polarization.

3. Empirical application

3.1 Data

In this section the expression defined is applied to empirical data from the European Community Household Panel (ECHP) for Spanish households in the period 1994-2001. The variable utilized to measure polarization is the annual equivalized net income. It is calculated using the modified OECD equivalence scale and to make comparisons we express the equivalized net income in constant euros at 2000 prices. The data were collected the year before making the survey, and for this reason we refer to 1993 and 2000. We have trimmed 1% of the upper and lower tails of the income distribution to avoid noise and bias in the estimation of polarization measures (see for instance Cowell and Victoria-Feser ,1996; Cowell and Victoria-Feser, 2002 and Weich et al. 2002).

Before computing polarization, we have to determinate the number of groups and their location. In line with most of the works about polarization (see for instance Gradín, 2000) we are going to compute polarization for two and three groups. The support interval of each income group is determined endogenously applying the K-means algorithm (McQueen, 1967). Therefore, we determine the size of the groups- two and three- in such a way that dispersion within groups is minimized. This algorithm assigns each individual to the group having the nearest centroid (mean). It is composed of four steps

1. Partition the items (individuals) into K initial clusters.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the position of the K centroids.
4. Repeat Step 2 and 3 until the centroids no longer move.

We run the k-means algorithm for two and three groups (the sizes and the means of the groups provided the algorithm are in Table 1). To interpret the results, we assume for the two-group case, that the first group is the less favored and the second is the more favored or privileged in terms of income. For the three-group case, we assume that the first group is formed by the less favored population in terms of income that is, poor and low middle classes. The second group is associated with the middle class and the third contains the upper middle class and the rich.

To establish comparisons, we calculate the measures of bi-polarization of Wolfson (1994, W henceforth), Esteban and Ray (1994), Zhang and Kanbur (2001) and Tsui and Wang (2001, TW henceforth), and the measures of Esteban and Ray (1994) and Zhang and Kanbur (2001) for three groups. The measure of ZK has been normalized (ZKN) following the approach that we have used to define I_{ia} in Section 2. The expressions of the listed measures and the normalization of the ZK measure are deferred to the appendix. We add the inequality indices of Gini and Theil for contrast and finally, we take 1993 equal to 100 for all the calculated measures to establish comparisons.

3.2 Results

During the considered period, the Spanish economy moved from an economic crisis to a path of stable and balance growth. Starting from a crisis situation in the year 1992, with high rates of inflation and unemployment, the gross domestic product, in constant prices, showed an annual negative percentage change in 1993. The economic downturn was serious but brief and after 1994 the economy started to grow with strong increases in the GDP, over 4%, from 1997 to 2000. In this favorable context in terms of economic growth, inequality diminished as it showed by the Gini and the Theil indices (see Table 3 and 4) whereas the conclusions about polarization depend on the utilized measure. Focusing on bi-polarization, the measures of ER, W and TW did not generate very different results from the Gini and the Theil indices[10] (Figures 2 and 4). All of them reached the maximum in 1996 and thereafter decreased like those of the inequality. However, the proposed measure and the normalized measure of ZK present an upward trend, although the latter less pronounced than the former, which involves an increase of bi-polarization from 1993 to 2000. More in detail and according to the polarization factors, it is observed that the identification-alienation factor, I_{ia}, presents a tendency almost horizontal, analogous to the normalized measure of ZK. The differences between both are due

[10] Zhang and Kanbur (2001) and Ravallion and Chen (1997) also found, using different data, a close correspondence between polarization and inequality measures.

to the use of the intra and inter-groups variance instead of the intra and inter-groups inequality. The factor with the greatest impact on polarization is I_m showing a pattern similar to the new measure. The upward trend of I_m is an effect of the evolution of the distribution of the size of the groups (see Table 1). For the overall period, it is observed that the group size gap decreased as a consequence of the transfers of population from the first to the second income group. These movements, supported by a favorable economic context especially after 1996, led to the two groups becoming closer in terms of mean income and size. As a result, the groups are more equally sized resulting in a rising I_m and hence polarization. Given that I_{ia} presents almost a horizontal trend, we can say that the increase of bi-polarization was mainly due to the shrinking of the group size gap, since there is more tension when the groups are equally sized than when there is a majority group.

For three groups, it is observed that PG-3 and ZKN has been fluctuating up and down without a tendency. Nonetheless, both measures are smoother than the bi-polar case. The expression of ER takes values between the Gini and the Theil coefficients showing a polarization pattern close to both inequality measures. Focusing on the polarization factors, it is observed that I_{ia} did not experience major changes since the mean income of the three groups (Table 2) increased with almost parallel trends, and V_B and V changed in similar proportions. The rise of I_{ia} with respect to the two-group case (Table 2) is due to the increase in the number of groups and it is counterbalanced by the factor I_g. Although there were movements of individuals between the three considered income classes (Table 1) they did not produce significant changes on I_m .

Summarizing our results, we have that bi-polarization increased in Spain from 1993 to 2000 according to the proposed measure and the normalized measure of Zhang and Kanbur (2001) whereas inequality decreased.

4. Final Remarks

In this paper a measure of polarization is proposed that is a non-decreasing function of three factors: the alienation-identification, the distribution of the share of powers of the groups and the number of groups. The first factor has been defined following the alienation-identification framework put forward by Esteban and Ray (1994) and Zhang and Kanbur (2001) but important differences with respect to these authors have been introduced. For statistical reasons and for the concept of polarization, we have used the ratio of the inter-group heterogeneity to intra-group homogeneity to compute alienation and identification. In this way the proposed measure differs from the measures defined by Wolfson (1994), Zhang and

Kanbur (2001) and Silber et al (2007) which use inequality indices to compute within group homogeneity and between groups heterogeneity. The second factor of polarization, the share of power of the groups, values the importance of the middle class as a factor of social stability. The smaller the size of the middle class the greater its contribution to polarization. As it has been shown, this factor jointly with the third one extends and corrects the approach to polarization of Zhang and Kanbur (2001) in the way indicated in the paper.

The empirical results show that bi-polarization increased from 1993 to 2000 according to the proposed measure, as well as the normalized measure of ZK. For three groups, PG-3 and ZKN has been fluctuating up and down without a tendency. The measures of ER for two and three groups, W and TW did not generate very different results from the Gini and the Theil indices. All of them reach the maximum in 1996 and thereafter decreased.

5. References

Cowell, F.A. and M-P Victoria-Feser (1996). Robustness Properties of Inequality Measures. Econometrica, 64, pp. 77-101.

Cowell, F.A. and M-P Victoria-Feser (2002). Welfare Rankings in the Presence of Contaminated Data. Econometrica, 70, pp. 1221-33.

Duclos, J. Y., Esteban J.M., and Ray D. (2004). Polarization: Concepts, Measurement, Estimation. Econometrica, 74, pp. 1337-1772.

Esteban J.M., and Ray D. (1994). On the Measurement of Polarization. Econometrica, 62, pp. 819-851.

Esteban J. M. (2002). Economic Polarization in the Mediterranean Basin. CREI, (10), Universitat Pompeu Fabra.

Fisher W. D. (1958). On Grouping for Maximum Homogeneity. Journal of the American Statistical Association, 53 (284), pp. 789-798.

Gradín C. (2000). Polarization by sub-populations in Spain, 1973-91. Review of Income and Wealth, 46 (4), pp. 457-474.

Gasparini L., M. Horenstein, E. Molina and S. Olivieri (2008). Polarización Económica, Instituciones y Conflicto. Ed. Uqbar.

Hermoso G. J. A. and Hernández B. A. (2000). Curso Básico de Estadística Descriptiva y Probabilidad. Ed. Némesis.

Hussain M.A. (2009). The sensitivity of income polarization. Journal of Economic Inequality, 7 (3), pp. 207-223.

McQueen, J.B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium of Mathemathical Statistics and Probability, 1 Berkeley, Calf.: University of California Press, pp. 281-297.

Ravallion, M. and S. Chen (1997). What can new survey data tell us about recent changes in distribution and poverty. World Bank Economic Review, 11(2), pp. 357-82, May.

Silber, J., J. Deutsch, and M. Hanoka, (2007) .On the Link between the Concepts of Kurtosis and Bi-polarization. Economics Bulletin, 4 (36), pp. 1-5.

Tsui, K and Wang, Y. (2000). Ordering and New Classes of Polarisation Indices. Journal of Public Economic Theory, 2 (3), pp. 349-363.

Weich, S., Lewis, G., Jenkins, S.P. (2002). Income Inequality and self-rated health in Britain. Journal of Epidemiology and Community Health, 56, pp. 436-441.

Wolfson, M. C. (1994). When Inequalities Diverge?. American Economic Review, 84, pp.353-58.

Zelli R. and Pittau M.G. (2007). Exploring patterns of income polarization using siZer. Journal of Quantitative Economics, 5, pp.101-111.

Zhang, X. and R. Kanbur (2001). What Difference Do Polarization Measure Make? An Application to China. Journal of Development Studies, 37, pp. 85-98.

Table 1. Sizes and means of the groups

		Group size		Mean		
2 Groups	p1	p2		m1	m2	
1993	0.778	0.222		6012.393	15713.541	
1994	0.757	0.243		6047.467	15193.744	
1995	0.762	0.238		6117.626	15503.326	
1996	0.769	0.231		6069.854	15637.629	
1997	0.758	0.242		6304.333	15859.751	
1998	0.742	0.258		6668.709	16347.487	
1999	0.762	0.238		7059.318	17184.293	
2000	0.728	0.272		7055.541	16791.949	
		Group Size			Mean	
3 Groups	p1	p2	p3	m1	m2	m3
1993	0.582	0.320	0.097	4967.975	10514.334	19573.951
1994	0.575	0.321	0.103	5080.348	10537.569	18931.971
1995	0.592	0.313	0.095	5200.181	10864.896	19698.730
1996	0.545	0.334	0.121	4883.829	10090.132	18560.789
1997	0.586	0.316	0.098	5359.674	11163.414	19893.768
1998	0.553	0.336	0.111	5570.834	11433.319	20294.328
1999	0.569	0.343	0.087	5913.181	12132.284	22202.825
2000	0.524	0.356	0.119	5833.408	11725.786	20654.393

Table 2. PG and factors of polarization for two and three groups

	PG -2	Iia 2	Im	Ig
1993	0.39656	0.66233	0.59873	1.00000
1994	0.43023	0.66263	0.64928	1.00000
1995	0.42172	0.66048	0.63851	1.00000
1996	0.41170	0.66307	0.62090	1.00000
1997	0.43317	0.66796	0.64851	1.00000
1998	0.45276	0.65957	0.68645	1.00000
1999	0.40862	0.64032	0.63816	1.00000
2000	0.47302	0.65957	0.71717	1.00000
	PG -3	Iia	Im	Ig
1993	0.35685	0.83090	0.64421	0.66667
1994	0.35961	0.83031	0.64966	0.66667
1995	0.35737	0.83070	0.64531	0.66667
1996	0.36182	0.82323	0.65927	0.66667
1997	0.35636	0.82630	0.64691	0.66667
1998	0.35889	0.83001	0.64859	0.66667
1999	0.33991	0.81921	0.62239	0.66667
2000	0.35622	0.83064	0.64327	0.66667

Table 3. Bi-polarization measures

	Gini	Theil	Iia 2	Im2	PG -2	ZKN-2	ER-2	TW	W
1993	100	100	100	100	100	100	100	100	100
1994	96.789	92.033	100.045	108.442	108.491	101.332	96.792	98.852	98.796
1995	97.296	93.083	99.721	106.644	106.346	101.241	96.538	98.747	98.226
1996	99.089	98.181	100.112	103.702	103.818	100.324	99.535	99.448	99.167
1997	96.559	91.259	100.849	108.313	109.233	102.170	96.305	98.588	98.251
1998	95.042	87.213	99.583	114.650	114.171	101.916	95.478	98.348	98.415
1999	93.774	84.851	96.677	106.585	103.043	99.187	93.067	96.769	94.387
2000	92.456	82.257	99.584	119.781	119.282	102.234	93.253	96.638	94.971

Table 4. Tri-polarization measures

	Gini	Theil	Iia 3	Im3	PG-3	ZKN-3	ER-3
1993	100	100	100	100	100	100	100
1994	96.789	92.033	99.930	100.847	100.776	100.576	94.414
1995	97.296	93.083	99.976	100.171	100.147	100.186	96.885
1996	99.089	98.181	99.077	102.339	101.394	100.608	99.486
1997	96.559	91.259	99.447	100.420	99.865	99.924	94.318
1998	95.042	87.213	99.893	100.681	100.573	101.569	87.743
1999	93.774	84.851	98.593	96.614	95.255	100.121	85.735
2000	92.456	82.257	99.969	99.855	99.824	102.162	85.288

Figure 1. Factors of polarization and PG-2

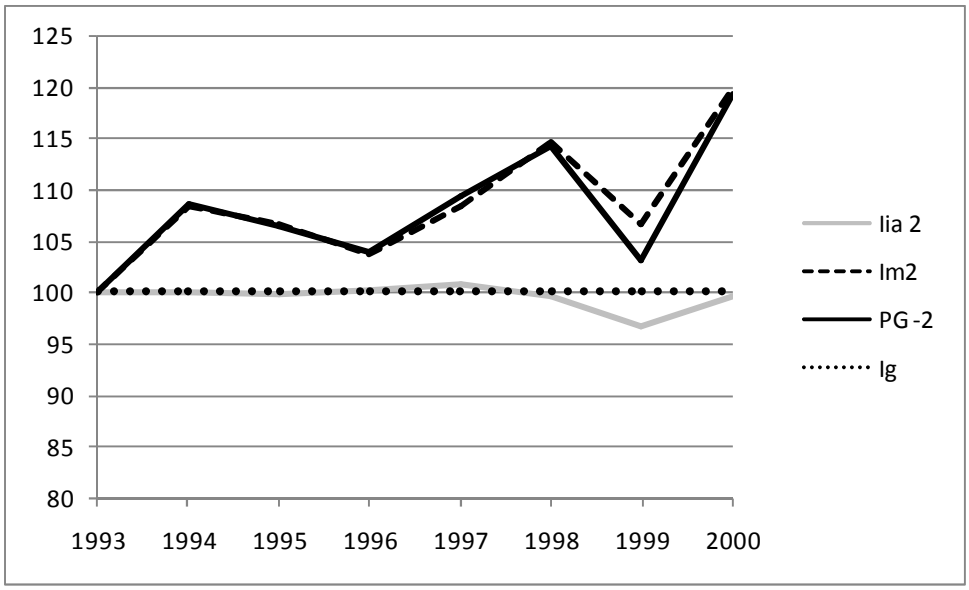


Figure 2. Bi-polarization measures and inequality indices

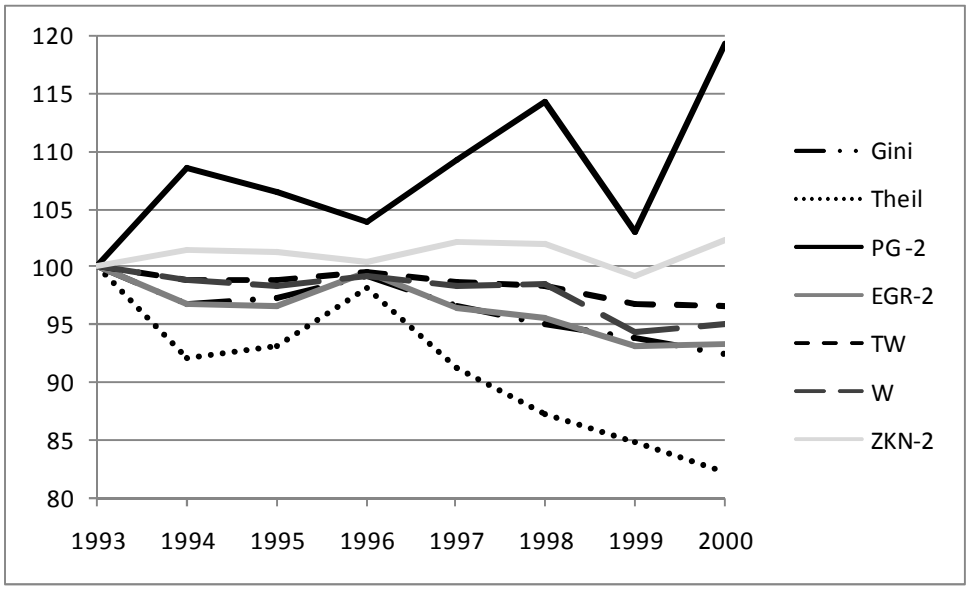


Figure 3. Factors of polarization and PG-3

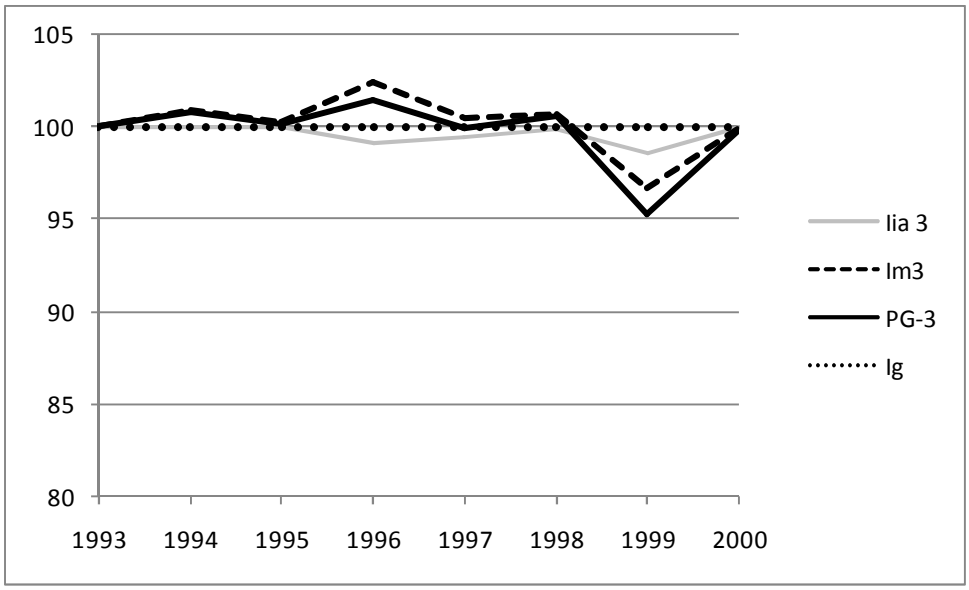
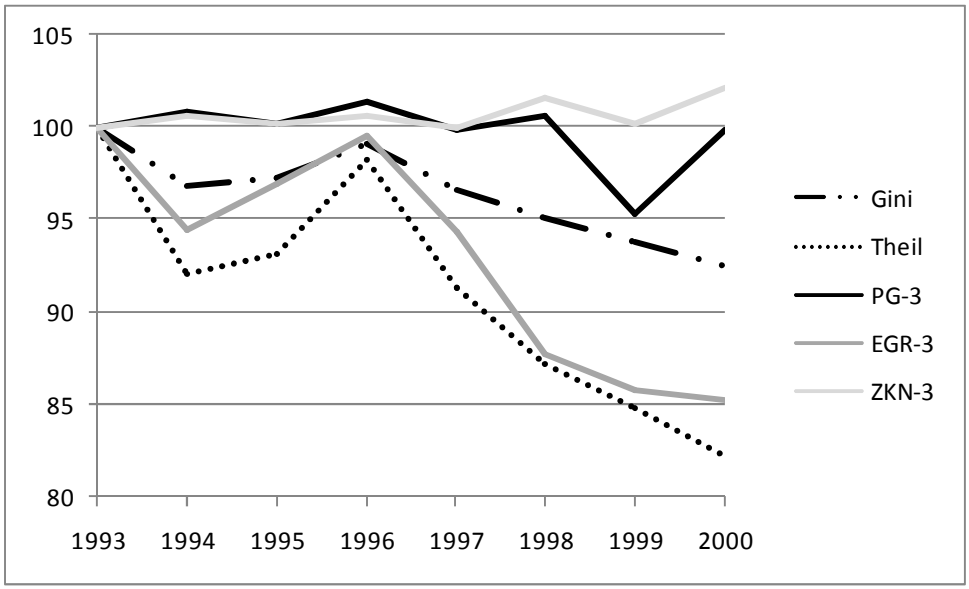


Figure 4. Tri-polarization measures and inequality indices



Appendix . Summary of polarization measures.

Esteban and Ray (1999) provided the following measure

$$ER(\alpha,\gamma)=\sum_{i=1}^{n}\sum_{i=1}^{n}\pi_i{}^{1+\alpha}\pi_j|\mu_i-\mu_j| \quad 1\leq\alpha\ \leq 1.6$$

where

$$\pi_i=\int_{y_{i-1}}^{y_i}f(y)dy$$

$$\mu_i=\frac{1}{\pi_i}\int_{y_{i-1}}^{y_i}yf(y)dy$$

represent the relative frequency and the conditional mean in group i for a density f of the logarithm of income respectively.

Zhang and Kanbur (2001) defined the following polarization measure

$$ZK=\frac{between-group\ inequality}{within-group\ inequality}$$

For the Theil index the above expression can be written as follows

$$ZK=\frac{T_B}{T_W}=\frac{\sum_{j=1}^{K}\frac{n_j}{N}\frac{\mu_j}{\mu}ln\left(\frac{\mu_j}{\mu}\right)}{\sum_{j=1}^{K}\frac{n_j}{N}\frac{\mu_j}{\mu}T_j}$$

where

$$T_j=\frac{1}{n_j}\sum_{j=1}^{K}\frac{y_j}{\mu_j}ln\left(\frac{y_j}{\mu_j}\right)$$

K is the number of groups, N is the total population, n_j is the population of the jth group, μ is the total sample mean, μ_j is the mean of the jth group and y_j is the jth income.

Observe that the ZK expression tends to infinite when the within-group inequality is equal to zero. This drawback of the measure can be corrected by following the approach that we use to define the index $I_{i\alpha}$. Proceeding in this way we obtain the normalized measure of Zhang and Kanbur which is given by

$$ZKN=1-\frac{T_W}{T}$$

where $T=T_W+T_B$.

Wolfson (1994) proposed the following measure of polarization based on the Lorenz curve

$$W = 4\frac{\mu}{m}\left[\frac{1}{2} - L\left(\frac{1}{2}\right) - \frac{G}{2}\right]$$

where μ is the mean, m is the median income, $L\left(\frac{1}{2}\right)$ is the Lorenz curve at the median income and G is the Gini index.

Tsui and Wang's (2000) expression which is connected to Wolfson's measure used two partial ordering axioms of "increased bipolarity" and "increased spread" to define the following class of indices:

$$TW = \frac{\theta}{N}\sum_{i=1}^{k} n_i \left|\frac{y_i - m}{m}\right|^r$$

where N is population total, n_i is the number of individuals that belong to group i k is the number of groups, y_i is the mean value in group i, m is the median income, θ is a positive constant and r takes values in the interval $(0,1)$.